

CIS 455/555 Homework2 Milestone 2 Testing Guidelines

The steps outlined below do not test all of the functionality that you are required to implement. This just gives you an idea of the basic things you should be able to do.

Before deploying the code, a few things to remember:

- 1 Please modify the web.xml file provided in HW2/conf to contain the mappings of all your servlets. If you are using a web.xml with **Tomcat/Jetty** for testing, please copy the servlet mappings `<servlet></servlet>` tags and the `<servlet-mapping></servlet-mapping>` class to the conf/web.xml before building and deploying your code.
- 2 We will be grading your assignments in the assigned grader's virtual machine. Hence, we will not accept relative paths to the Berkeley DB store in the web.xml, as it could be hard-coded for a specific location in your VM. So your web.xml should have lines like:

```
<context-param>
<param-name>BDBstore</param-name>
<param-value>/home/cis455/export/db</param-value>
</context-param>
```

- 3 Please communicate the path to the Berkeley database store that you will use with your crawler in your **README** file.
- 4 Please refrain from hard coding the database path in your java code.

Part 1: Compile and deploy

Compile:

The system you build for assignment 2 has two parts, namely the crawler and the web interface. To compile the two parts and execute them we will use the ant build script:

In the VM, open a new terminal.

Switch to the HW2 folder in your workspace directory using the command:

```
cis455@vm:~$ cd /home/cis455/workspace/HW2
cis455@vm:~/workspace/HW2$
```

Then execute the new build.xml file with the following command:

```
cis455@vm:~/workspace/HW2$ ant
```

The command builds your crawler into a `crawler.jar` file, and the web application with your servlets into a `servlet.war` file.

Deploy:

For Apache Tomcat 7:

Step 1: Copy the `servlet.war` file to `/var/lib/webapps/`

Step 2: To start tomcat, type `sudo service tomcat7 start`

Step 3: Open a browser, type `localhost:8080/servlet/XXXX`(url you defined in the `web.xml`) to access the user interface.

Step 4: To stop tomcat, type `sudo service tomcat7 stop`

For Jetty:

Step 1: Copy the `servlet.war` file to `directory_of_jetty/webapps/`

Step 2: Change to the directory where you unpacked jetty, using the command `cd directory_of_jetty`

Step 3: Use `java -jar start.jar` to start the server(every time you make modifications to the web application you need to restart the server).

Step 4: Open a browser, type `localhost:8080/servlet/XXXX`(url you defined in the `web.xml`) to access the user interface.

Part 2: Testing the system:

Creating a user and channel:

After compiling and deploying the web application, you can access the user interface on browser and test the application by executing the following steps:

1. Create a user.
2. Log in to the web interface with username and password.
3. Create a channel by supplying `its name, xpath expression(s) and a link` (local or web) to the XSL style sheet for that channel.

Running the crawler:

After you have successfully compiled the project, use the following command under the project root directory (/home/cis455/workspace/HW2) to start the crawler:

```
java -cp crawler.jar:target/WEB-INF/lib/servlet-api.jar:target/WEB-INF/lib/jtidy-  
r938.jar:target/WEB-INF/lib/je-5.0.34.jar edu.upenn.cis455.crawler.XPathCrawler URL  
DBDPath MaxSize [MaxNumOfFiles - if you have coded for this]
```

Note:

1. Put every external library(.jar file) you need into the classpath
2. The crawler will take 4 parameters: the seed url, the directory path to Berkeley DB (ensure it is the same as the one declared in web.xml), the maximum size of the crawled file (in Megabytes !) and the maximum number of files the crawler will crawl (optional)
- 3 The directory of Berkeley DB here should be the same directory as the value of "DBDstore" in the web.xml

For example:

```
java -cp crawler.jar:target/WEB-INF/lib/servlet-api.jar:target/WEB-INF/lib/jtidy-  
r938.jar:target/WEB-INF/lib/je-5.0.34.jar edu.upenn.cis455.crawler.XPathCrawler  
http://crawltest.cis.upenn.edu/ /home/cis455/export/db 10 10
```

Display channels

Log in to web interface again, and use the display channel functionality to see if the xpath matching and XSL works fine.